University of Southern Denmark
IMADA
Arthur Zimek
Jonatan Møller Gøttcke, Jonas Herskind Sejr

DM566/DM868/DM870: Data Mining and Machine Learning

Spring term 2019

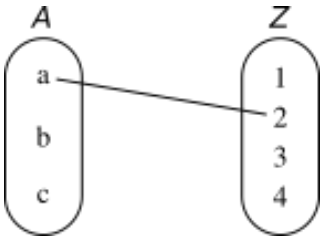
Exercise 2: Relations, Itemsets and Association Rules

Exercise 2-1 Sets, Relations, Functions – Visualization

Consider the sets $A = \{a, b, c\}$ and $Z = \{1, 2, 3, 4\}$ and some binary relation over them.

If for example the elements $a \in A$ and $2 \in Z$ are in relation R could we write: $aR2$ or $(a, 2) \in R$.

As a graphical visualization we can draw the two sets and a line connecting a and 2:



Given such a visualization, the mathematical definitions basically tell us which lines to draw.

Give such a visualization for:

(a) The Cartesian product $A \times Z$

(b) A binary relation over A and Z, that is not a function.

(c) A non-total function $A \to Z$

(d) A total function $A \to Z$

(e) We called the case c) *non-total*. Could we have called it *partial*, or would that make a difference?

Exercise 2-2 Itemsets and Association Rules

Given a set of transactions T according to the following table:

Set of transactions T

Transaction ID	items in basket
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies }
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

(a) What are the support and the confidence of $\{\text{Milk}\} \Rightarrow \{\text{Diapers}\}$?

(b) What are the support and the confidence of $\{\text{Diapers}\} \Rightarrow \{\text{Milk}\}$?

(c) Write an expression for the maximum number of size-3 itemsets that can be derived from this data set.

(d) What is the maximum number of association rules that can be extracted from this data (including rules, that have zero support)?

(e) What is the maximum size of frequent itemsets that can be extracted (assuming $\sigma > 0$)?

(f) Find an itemset (of size 2 or larger) that has the largest support.

(g) Find a pair of items, a and b, such that the rules $\{a\} \Rightarrow \{b\}$ and $\{b\} \Rightarrow \{a\}$ have the same confidence.